差分隐私模型的启发式隐私参数设置策略 *

欧阳佳[1]，肖政宏[1]，刘少鹏[1]，印　鉴[2]，林丕源[3]

(1. 广东技术师范学院 计算机科学学院，广州 510665; 2. 中山大学 数据科学与计算机学院，广州 510275; 3. 华南农业大学 数学与信息学院，广州 510642)

摘　要：差分隐私模型是一种强隐私模型，用隐私参数 ε 度量隐私保护程度及噪声量，近年来成为隐私保护领域的研究热点。但是隐私参数 ε 的设置只能依赖于实验或专业人士经验，限制了差分隐私模型的使用与推广。针对这个问题，基于(ρ1，ρ2)-隐私模型提出一种启发式的隐私参数 ε 设置策略（limit privacy breaches in differential privacy，LPBDP），分析隐私参数 ε 与(ρ1，ρ2)的内在联系，实现噪声量的添加由(ρ1，ρ2)决定。LPBDP 通过如下启发式原则设置隐私参数 ε：如果攻击者关于目标受害者的先验概率小于阈值 ρ1，攻击者得到差分隐私查询策略返回的加噪结果后，关于目标受害者的后验概率必须小于阈值 ρ2。实验表明 LPBDP 能够更直观地设置隐私参数 ε 以满足差分隐私约束。

关键词：隐私保护；差分隐私；隐私参数；隐私泄露

中图分类号：TP309.2　　doi: 10.3969/j.issn.1001-3695.2017.06.0649

Heuristic privacy parameter setting strategy for differential privacy model

Ouyang Jia[1], Xiao Zhenghong[1], Liu Shaoeng[1], Yin Jian[2], Lin Piyuan[3]

(1. *College of Computer Science Guangdong Polytechnic Normal University*, *Guangzhou* 510665, *China*; 2. *School of Data & Computer Science*, *Sun Yat-sen University*, *Guangzhou* 510275, *China*; 3. *College of Mathematics & Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: The differential privacy model is a kind of strong privacy model, which uses the privacy parameter ε to measure the degree of privacy protection and the amount of noise. In recent years, the privacy model has become a hotspot in the field of privacy protection. However, the setting of the privacy parameter ε can only depend on the experience of the lab or the professional experience, limiting the adoption and popularize of the differential privacy model. Aiming at this problem, a kind of heuristic privacy parameter ε setting strategy（limit privacy breaches in differential privacy, LPBDP）is proposed based on the (ρ1, ρ2) -privacy model. The intrinsic relationship between the privacy parameter ε and (ρ1, ρ2) is analyzed, and the addition of the noise quantity is determined by the parameters (ρ1, ρ2) . LPBDP sets the privacy parameter ε by the following heuristic principle: If the attacker's prior probability of the target victim is less than the threshold ρ1, then, the attacker's posterior probability of the victim of the target must be less than threshold ρ2. Experiments show that LPBDP can more visually set the privacy parameter ε to meet the differential privacy constraints.

Key Words: privacy-preserving; differential privacy; privacy parameter; privacy breaches

0　引言

随着计算机科学、网络以及存储技术的发展，人类社会收集、存储的数据已经达到了前所未有的程度，数据的爆炸式增长又促进了数据挖掘的巨大发展，数据挖掘技术已经成功应用于社会的各行各业，如：医疗、社交网络、在线搜索等领域。由于对原始数据的访问是数据挖掘的前提，但原始数据中往往包含个人的隐私信息，因此随之而来的是个人对隐私保护越来越关注。目前，数据挖掘领域中一个重要的研究方向是准确得到知识的同时保证数据与个人隐私的安全。

隐私保护数据挖掘的出现就是为了解决上述数据挖掘所带来的隐私担忧问题。隐私保护数据挖掘的目的是能成功构建各

基金项目：国家自然科学基金项目（61702119）；广东省教育厅青年创新人才项目（自科）（57/572020507）；广州市科技计划项目（201607010152）；广东省省级科技计划项目（2016A010101029）

作者简介：欧阳佳（1986-），男，湖南新化人，讲师，博士研究生，主要研究方向为机器学习、隐私保护、差分隐私（ouyangjia1@163.com）；肖政宏（1965-），男，教授，博士研究生主要研究方向为大数据理论与技术、机器学习、网络信息安全；刘少鹏（1984-），男，讲师，博士研究生，主要研究方向为机器学习与深度学习、文本挖掘；印鉴（1968-），男，教授，博导，博士研究生，主要研究方向为信息技术、数据挖掘、电子商务；林丕源（1963-），男，教授，硕士研究生，主要研究方向为生物信息学、信息安全、智能计算与数据挖掘．

种有效的数据挖掘模型而不会泄露输入的原始数据[1~4]。具体来说，隐私保护数据挖掘需要解决如下两个关键问题：a) 如何在数据挖掘的过程中保护个人隐私；b) 如何确保数据或结果的效用性。目前，隐私保护数据挖掘主要集中于隐私准则的设计以及同时满足上述两个关键点的算法。

ε-差分隐私模型[5~8]不对攻击者的背景知识作任何假设，是一种隐私保护力度非常强的隐私模型，它的基本思想是发布对敏感数据的分析结果之前，添加少量噪声以满足差分隐私要求，噪声量由分析函数或分析过程的敏感度以及隐私参数 ε 共同决定，与具体的数据库类型及其大小无关。

隐私参数 ε 是差分隐私模型的重要参数，用于决定噪声的添加量以及度量隐私保护的程度。从拉普拉斯机制与指数机制中可以看出，ε 越大，添加的噪声越少，相反，ε 越小，添加的噪声越多。但差分隐私模型在决定添加噪声量的多少时存在两个方面的问题：第一个问题是隐私参数 ε 仅仅限制了个体记录对结果的影响，而不是限制个人泄露了多少信息[9]，将导致攻击者在获得随机结果后很容易识别个人的敏感信息；第二个问题是隐私参数 ε 的设置只能依赖于实验或专业人士的经验，没有更加直观的启发式参数设置方法。

针对上述两个问题，本文的主要贡献如下：

a）为限制差分隐私模型中个人信息的泄露，基于(ρ_1, ρ_2)-隐私模型的思想，提出一种新的攻击模型；

b）找出隐私参数 ε 与(ρ_1, ρ_2)之间的关系，提出了一种启发式隐私参数设置策略。

1 相关工作

针对恶意攻击者的攻击手段，研究者已提出众多优秀的隐私模型，如针对链接攻击的 *k*-匿名模型[11]；针对属性攻击的 *l*-多样化模型[12]。这些模型的基本思想都是基于数据的匿名分组，匿名过程将整个数据集划分为多个等价类，每个等价类中至少包含 *k* 条记录，恶意攻击者识别个体的概率最多为 1/*k*。

差分隐私(differential privacy)[6]是一种完全独立于攻击者背景知识和计算能力的强隐私概念，近年来已成为研究热点。它假设攻击者拥有任意的背景知识，无论特定个体记录是否在数据集中，对该数据集的任意计算分析或查询的结果在形式上不可区分。差分隐私随机算法对任意两个邻近数据集进行操作，得到的结果几乎是一致的。形式化来说，已知 D 为任意数据集，设与 D 只相差一条记录的近邻数据集为 D'，差分隐私要求任意算法对 D 与 D' 得到相同结果的概率的比值有一个常数上界。差分隐私模型体系中最基本的模型是 ε-差分隐私模型，其定义如下：

ε-差分隐私(differential privacy)[6]：随机算法 Π 满足 ε-差分隐私约束，如果任意的两个邻近数据集 D 和 D'，$|D\Delta D'|=1$，对于所有输出数据集 O，下列不等式成立：

$$\Pr\left[\Pi(D)=O\right]\le e^{\varepsilon}\Pr\left[\Pi(D')=O\right] \tag{1}$$

其中：$|D\Delta D'|=1$ 表示数据集 D 和 D' 只有一条记录不同，ε-差分隐私保证 D 中任意一条记录的改变对算法 Π 输出的影响都不会过于明显。拉普拉斯机制与指数机制是满足差分隐私约束的两种标准方法，它们均依赖于函数的全局敏感度。

Lee 等人[9]提出了一种 ρ-差分可识别(differential identifiability)的概念，ρ-差分可识别提供与 ε-差分隐私模型一样的隐私保护力度，它的优点是参数 ρ 限定了每个个体对分析结果贡献的概率估计。ρ-差分可识别通过限定攻击者对个体的后验概率将隐私参数 ε 与 ρ 联系在一起，数据挖掘者或数据发布者可以基于 ρ-差分可识别设置隐私参数 ε，添加的噪声限定攻击者在获得分析结果后推断目标受害者敏感值的概率不高于 ρ。隐私机制 M 满足 ρ-差分可识别约束，需要隐私参数 ε 与 ρ 满足如下关系：

$$\varepsilon=\ln\frac{(m-1)\rho}{1-\rho} \tag{2}$$

其中：U 为所有可能的值以及 $m=|U|$，在假设 U 中所有的 i 的先验概率 $\Pr[D=D'\cup\{i\}]$ 都相等以及 $|U|$ 已知的情况下，ρ-差分可识别与差分隐私之间存在一种联系，即：任意的 ρ-差分可识别隐私机制都满足 $\ln(m-1)\rho/(1-\rho)$-差分隐私约束。从式(2)中得出，ρ-差分可识别依赖于个体的先验分布，并假设预先知道所有可能的值 U 以及$|U|$。然而现实中，个体的先验分布一般都是不相等的，甚至根本无法预先获得先验分布，并且不一定完全知道 U 的值。因此本文基于(ρ_1, ρ_2)-隐私模型提出一种独立于这种先验分布的隐私参数设置策略。

2 基于(ρ1, ρ2)-隐私模型的隐私参数设置

2.1 隐私模型

文献[10]首次提出(ρ_1, ρ_2)-隐私模型的概念，它的定义为：当随机变量 X 的值 x 的先验概率 $\Pr[X=x]\le\rho_1$ 时，通过隐私策略 M 得到扰乱结果 $R\in Range(M_f(D))$ 后，x 的后验概率更新为 $\Pr[X=x\,|\,M_f(D)=R]\le\rho_2$，则称 x 满足(ρ_1, ρ_2)-隐私模型约束。其中，$0<\rho_1<\rho_2<1$ 并且 $\Pr[M_f(D)=R]>0$。从定义上看，(ρ_1, ρ_2)-隐私模型并不依赖于先验概率，它意味着先验概率不超过 ρ_1，则后验概率必须小于 ρ_2，ρ_1 与 ρ_2 可以自定义不依赖于任何背景知识。

2.2 攻击模型

假设攻击者的背景知识包含所有可能的值 U 以及数据库 D 中除了第 n 个元组以外其他所有元组的信息，也就是 D'。另外攻击者还知道隐私机制 M 的所有细节以及添加噪声所服从的概率密度函数。攻击者为了推断第 n 个元组的值，在得到隐私机制 M 返回的结果前攻击者以相等的概率 $1/U$ 猜测 U 中所有的值都有可能为第 n 个元组的值，用户提交查询 f 给隐私机制 M，得到扰乱的结果为：$R=M_f(D)$，则攻击者猜测第 n 个元组的值为 i 的概率为

$$\Gamma(i)=\Pr\left[w=D\,\middle|\,M_f(D)=R\right] \tag{3}$$

如果$\Gamma(i) > \rho_2$，则隐私泄露了。

2.3　攻击模型举例

下面通过一个例子描述上述攻击模型的过程，尽管查询机制M满足差分隐私约束，但攻击者依然可以以很高的后验概率猜测个体的值。令f为求平均值的查询函数，给定数据集$D=\{1,2,3,10\}$，D中的值都来自$U=\{1,2,3,5,10\}$，假设攻击者已经知道$D'=\{1,2,3\}$，想推断第4个值，由于第4个值可能为1,2,3,5,10，得知f的敏感度为：$16/4-7/4=9/4$。设差分隐私参数$\varepsilon=2$，且攻击者提交一个求平均值的请求后得到的返回值为：$R=5.041$，缺失的值为U中的其中一个，攻击者计算后验概率$\Pr[X=x \mid M_f(D)=R]$，如表1所示。

表1　攻击者的猜测值

猜测值	猜测数据集	真实均值	添加噪声	Pr[$M_f(D_i)$]=5.401	后验概率
1	1,2,3,**1**	7/4	3.291	0.0238	0.0751
2	1,2,3,**2**	8/4	3.401	0.0216	0.0682
3	1,2,3,**3**	9/4	2.791	0.0372	0.1174
5	1,2,3,**5**	11/4	2.291	0.0580	0.1831
10	**1,2,3,10**	**16/4**	**1.041**	**0.1762**	**0.5562**

以可能的值10为例给出后验概率的计算过程。

$$\begin{aligned}&\Pr\left[X=10 \middle| M_f(D_i)=5.401\right]\\&=\frac{\Pr[X=10]\cdot\Pr\left[M_f(D_i)=5.401 \middle| X=10\right]}{\sum_{i\in U}\Pr[X=i]\cdot\Pr\left[M_f(D_i)=5.401\right]}\\&=\frac{\Pr[X=10]\cdot\Pr\left[M_f(D_{10})=5.401\right]}{\sum_{i\in U}\Pr[X=i]\cdot\Pr\left[M_f(D_i)=5.401\right]}\end{aligned} \tag{4}$$

其中：$D_i=\{D'\cup i\}$。首先基于返回的值5.401，计算概率$\Pr\left[M_f(D_{10})=5.401\right]$，因为$mean(D_{10})=4$，差分隐私机制$M$给真实值添加的噪声量为：$R-mean(D_{10})=1.041$，可以求出$\lambda=\frac{\Delta}{\varepsilon}=\frac{9}{8}=1.125$，则

$$\begin{aligned}&\Pr\left[M_f(D_{10})=5.401\right]\\&=\frac{1}{2\cdot 1.125}\cdot e^{-\frac{|1.041|}{1.125}}\\&=0.1762\end{aligned} \tag{5}$$

假设U中值的先验概率为$\rho_1=0.2$，则

$$\begin{aligned}&\Pr\left[X=10 \middle| M_f(D_i)=5.401\right]\\&=\frac{\Pr\left[M_f(D_{10})=5.401\right]}{\sum_{i\in U}\Pr\left[M_f(D_i)=5.401\right]}\\&=0.5562\end{aligned} \tag{6}$$

如果$\rho_2=0.5$，则差分隐私机制M不满足(ρ_1,ρ_2)-隐私约束。

2.4　LPBDP的设计与实现

本文基于(ρ_1,ρ_2)-隐私模型提出的隐私参数设置策略LPBDP(limit privacy breaches in differential privacy)的基本思想是设置隐私参数ε使得差分隐私机制满足(ρ_1,ρ_2)-隐私约束，找出差分隐私参数ε与(ρ_1,ρ_2)之间的关系，使得ε的设置不再依赖于经验或实验，而是可以根据(ρ_1,ρ_2)进行启发式设置。

为满足(ρ_1,ρ_2)-隐私约束，文献[10]提出一种增幅(amplification)方法。该方法的定义如下：隐私机制M对于所有的结果$R\in M_f(D)$最多是γ-增幅的，如果式(7)成立：

$$\forall D_i, D_j: \frac{\Pr\left[D_i\to R\right]}{\Pr\left[D_j\to R\right]}\le\gamma \tag{7}$$

其中：$\gamma\ge 1$，$D_i=\{D'\cup i \mid i\in U\}, D_j=\{D'\cup j \mid j\in U\}$。

如果隐私机制M返回的结果是R，那么任意一个数据集$D=\{D'\cup i \mid i\in U\}$都有可能返回$R$。因此，基于拉普拉斯机制得到如下等式：

$$\begin{aligned}\frac{p\left[D_i\to R\right]}{p\left[D_j\to R\right]}&=\frac{\frac{1}{2\lambda}\cdot e^{-\frac{|R-f(D_i)|}{\lambda}}}{\frac{1}{2\lambda}\cdot e^{-\frac{|R-f(D_j)|}{\lambda}}}\\&=e^{\frac{|R-f(D_j)|-|R-f(D_i)|}{\lambda}}\end{aligned} \tag{8}$$

因为$\left|f(D_i)-f(D_j)\right|\le\Delta$，应用三角不等式得到

$$\frac{p\left[D_i\to R\right]}{p\left[D_j\to R\right]}\le e^{\frac{|f(D_i)-f(D_j)|}{\lambda}}\le e^{\frac{\Delta}{\lambda}} \tag{9}$$

下面的定理[10]给出了γ-增幅与(ρ_1,ρ_2)-隐私模型之间的关系。

定理1　如果ε-差分隐私机制M对于所有的响应值R都满足γ-增幅，其中$\gamma\le\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}$，则$M$必定满足$(\rho_1,\rho_2)$-隐私约束。

基于定理1，可以找出差分隐私参数ε与(ρ_1,ρ_2)之间的关系。根据式(9)得到：

$$\frac{p\left[D_i\to R\right]}{p\left[D_j\to R\right]}\le e^{\frac{\Delta}{\lambda}}\le\gamma\le\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2} \tag{10}$$

因为$0<\rho_1<\rho_2<1$，两边同时取自然对数得到

$$\frac{\Delta}{\lambda}\le\ln\left(\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}\right) \tag{11}$$

$$\lambda\ge\frac{\Delta}{\ln\left(\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}\right)} \tag{12}$$

由此得到一个重要的结果，对于任意的攻击者，如果设置差分隐私参数$\varepsilon=\ln\left(\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}\right)$，则拉普拉斯分布的参数：

$$\lambda = \Delta \Big/ \ln\left(\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}\right) \tag{13}$$

那么差分隐私机制 M 满足 (ρ_1,ρ_2)-隐私约束。又由于拉普拉斯分布要求 $\lambda>0$，则有

$$\ln\left(\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}\right)>0 \tag{14}$$

$$\frac{\rho_2}{\rho_1}\cdot\frac{1-\rho_1}{1-\rho_2}>1 \tag{15}$$

最后得到：$\rho_2>\rho_1$，意味着保护数据库中个体的隐私的后验概率 ρ_2 必须超过它的先验概率 ρ_1，否则没有任何意义。这个结论显然是符合实际的，如果对个体实施隐私保护的概率不能超过它的先验概率，那就失去了保护的意义。

注意到，如果设置 $\rho_1=1/m$，其中 $m=|U|$，就能得到：

$$\lambda \ge \Delta \Big/ \ln\frac{(m-1)\rho_2}{1-\rho_2} \tag{16}$$

意味着 ρ-差分可识别只是本文提出方法 LPBDP 的一个特例。

3 实验结果与分析

本节首先对比 LPBDP 与 ρ-差分可识别的区别与联系，如表 2 所示；其次通过实验分析了 LPBDP 的启发性和语义性。

一方面，从表 2 中可以看出，本文提出的 LPBDP 方法同样满足差分隐私要求，且比 ρ-差分可识别在先验知识上更有优势，LPBDP 基本上不需要假设任何先验知识，具有更好的适应性。

表 2 LPBDP 与 ρ-差分可识别的区别与联系

	区别			联系
	攻击模型	先验知识	差分隐私	关联
LPBDP	先验概率大于 ρ_1 后验概率小于 ρ_2	ρ_1（根据实际需求自定义设置）	满足	当 $\rho_1=1/m$，其中 $m=\lvert U\rvert$ 时，ρ-差分可识别为 LPBDP 的特例
ρ-差分可识别	后验概率小于 ρ	数据集 U 中存在的每个值； 数据集 U 的大小 $\lvert U\rvert$； 每个值先验概率相等且都为随机猜测概率 $1/\lvert U\rvert$。	满足	

另一方面，通过实验分析 LPBDP 的实际应用，为了与 ρ-差分可识别进行比较，实验中采用同样的聚集查询函数，即求平均值：mean，实验数据为来自 UCI 的 Adult 数据库，包含 48,842 条记录，共有 14 个属性，其中 9 个分类属性，5 个数值型属性。在本文中只用到其中 3 个数值型属性。表 3 描述了 Adult 数据库的特点。

表 3 Adult 数据库

属性	最大值	最小值	敏感度	随机猜测概率
age(AG)	90	17	0.0015	0.0137
education-num(EN)	16	1	0.0031	0.0101
hourse-per-week(HW)	99	1	0.0020	0.0625

为了决定所添加噪声的拉普拉斯分布函数，必须求平均值函数的敏感度：Δf。例如，假设攻击者知道数据库中除一条记录外其他所有记录的年龄，那么攻击猜测值的范围为 1~99。所以，函数的敏感度为

$$\Delta f = \left|f(D_{90})-f(D_{17})\right| = \frac{90-17}{48842} = 0.0015 \tag{17}$$

攻击者随机猜测的概率为 $1/|U|$，如表 3 中 RG（random guess）所示。

LPBDP 表明，噪声添加量不仅受先验概率影响，也受后验概率影响。本文首先通过实验验证了添加的噪声量受先验概率影响的情况，设置后验概率为：$\rho_2=50\%$。要求差分隐私机制满足 $(\rho_1,0.5)$-隐私要求，其中 $\rho_1=1\%\ldots10\%$。如图 1 所示，噪声添加量 (λ) 随着先验概率的增大而增大。意味着 $\rho_2-\rho_1$ 越小，所需的噪声量越多。ρ-差分可识别中将所有值的先验概率都设置为随机猜测概率 $1/|U|$，对于有的先验概率大于随机猜测概率的值，所添加的噪声量不能满足 (ρ_1,ρ_2)-隐私模型约束。

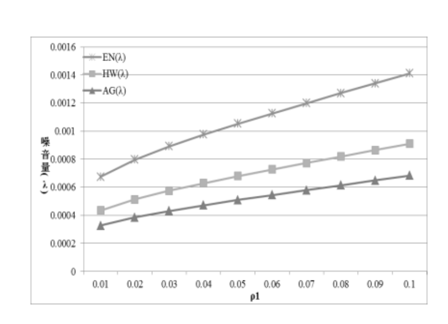
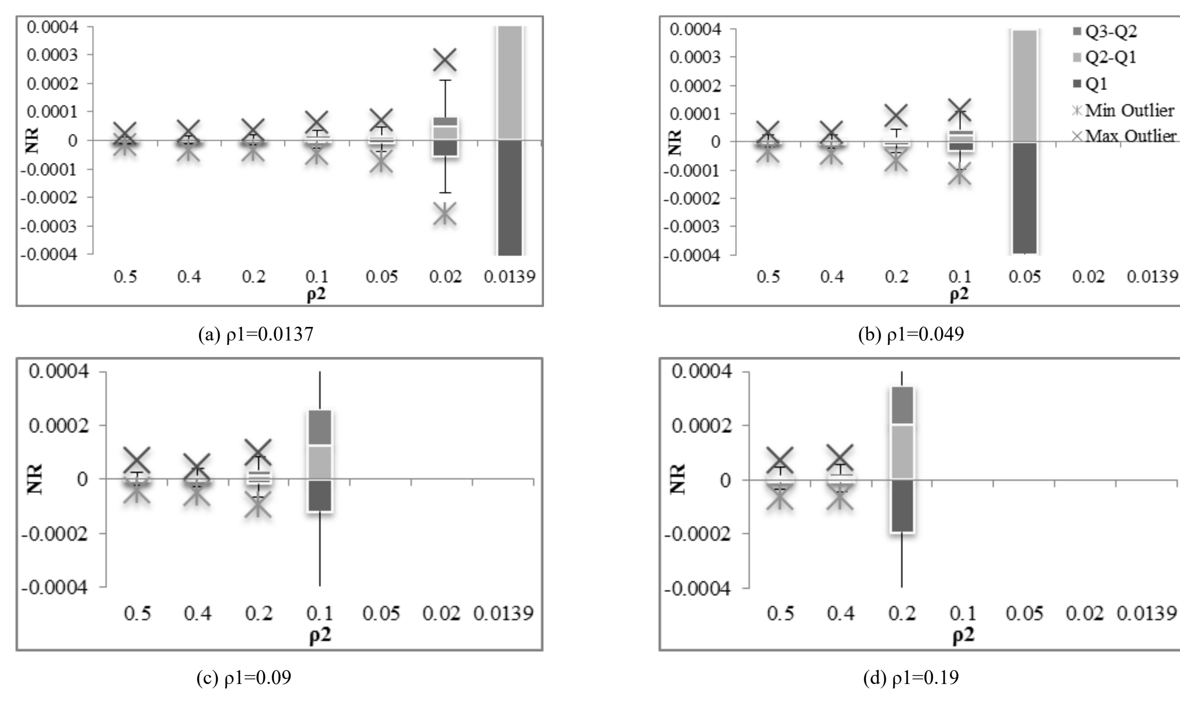


图 1 先验概率对噪声添加量的影响

为了验证 LPBDP 的实用性，对 4 个属性分别提交了 1000 次求平均值的查询请求，图 2~4 给出了 ρ_1,ρ_2 对噪声率的影响，噪声率的计算为

$$\text{Noise ratio}(\text{NR}) = \frac{R-f(D)}{U_{range}} \tag{18}$$

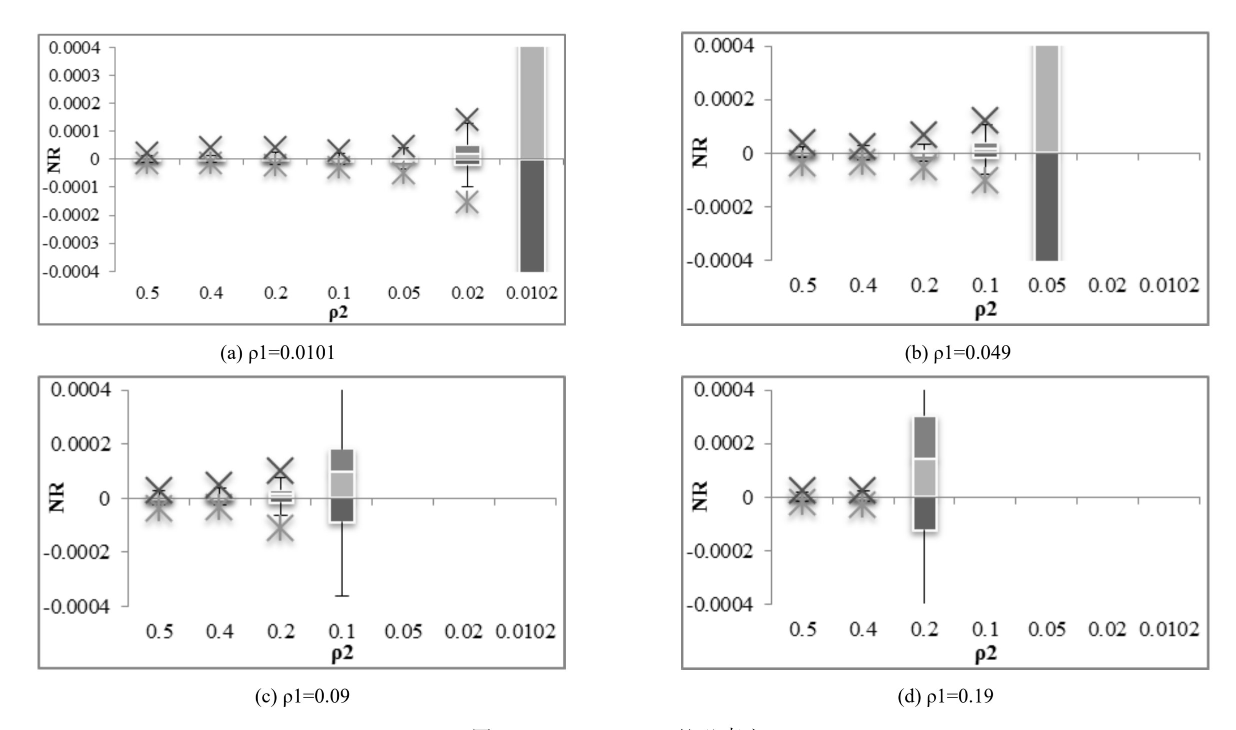
其中 R 为扰乱后的查询结果，$U_{range}=\max-\min$ 是每个属性域上值的区间。Q1 为第一个四分位数，Q3 为第 3 四分位数，Q3-Q1 为分位数极差。图 3~4 表明，所有的响应值都集中在真实值的附近。当固定 ρ_1 时，增大 ρ_2，需要更多的噪声以满足 (ρ_1,ρ_2)-隐私约束，这与 ρ-差分可识别是相同的。当固定 $\rho_2=0.2$，从图 3、4 中可以看出（从左至右），当 ρ_1 增大时，λ 随之变大，因此所需要的噪声也增多。



(a) ρ1=0.0137　　(b) ρ1=0.049

(c) ρ1=0.09　　(d) ρ1=0.19

图 2　Age 的噪声率



(a) ρ1=0.0101　　(b) ρ1=0.049

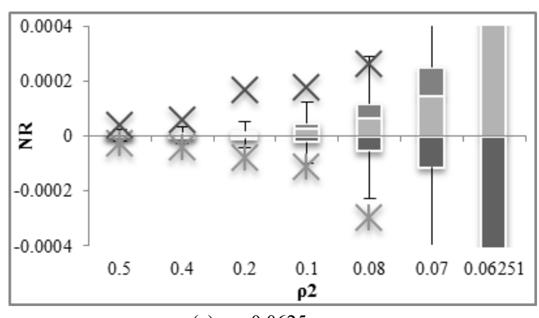
(c) ρ1=0.09　　(d) ρ1=0.19

图 3　Hours-per-week 的噪声率

图 5 研究了隐私参数 ε 对差分隐私的影响。图 5 表明 LPBDP 能达到 ε -差分隐私一样的效果。但是对于 ε -差分隐私而言，如何设置 ε 是个大问题，大多通过经验或实验设置。而 LPBDP 设置隐私参数具有更好的语义，必须满足 (ρ_1,ρ_2) -隐私约束。

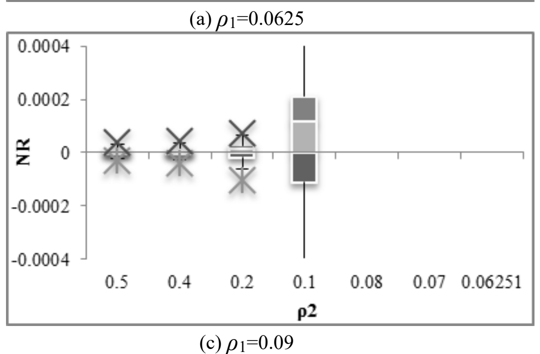
4　结束语

针对差分模型中的隐私参数设置问题，以往的设置方法主要基于实验或相关专家的经验，本文提出了一种启发式的差分隐私参数设置策略。ρ -差分可识别是另一种差分隐私参数的设置策略，然而该方法依赖于如下两个假设：(1) 知道每个值的先验概率，并假设预先知道所有可能的值 U 以及 $|U|$；(2) 所有可能值的先验概率都是相等的。然而，部分应用场景无法满足上述两个假设条件，本文提出的方法弥补了这一缺陷，基于 (ρ_1,ρ_2) -隐私模型提出一种新的隐私参数设置策略 LPBDP，该策略的优势在于 (ρ_1,ρ_2) -隐私模型并不依赖于先验概率且不需要知道 $|U|$，且 LPBDP 同样满足差分隐私约束。

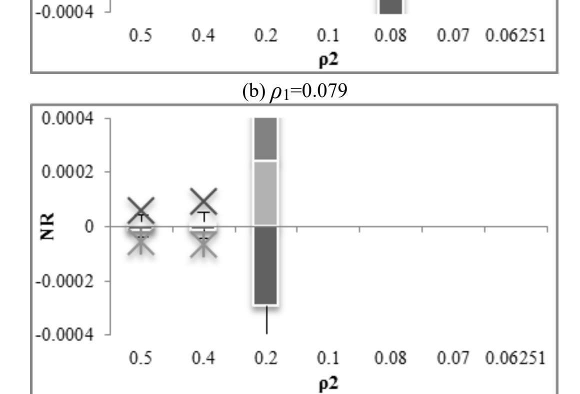


(a) ρ_1=0.0625

(b) ρ_1=0.079

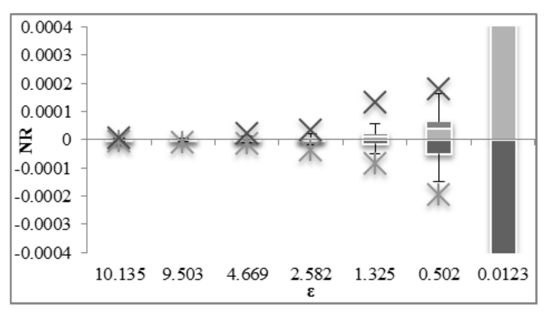


(c) ρ_1=0.09

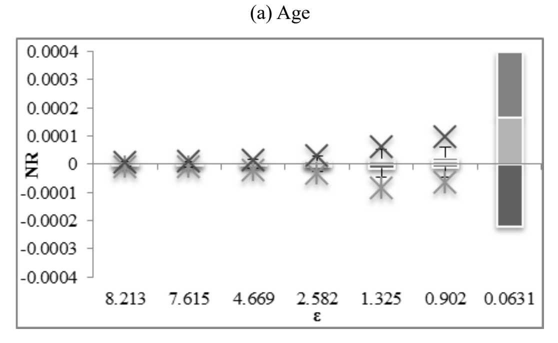


(d) ρ_1=0.19

图 4 Education Number 的噪声率



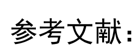
(a) Age



(b) Hours-per-week

(c) Education Number

图 5 差分隐私的噪声率



参考文献：

[1] 欧阳佳，印鉴，刘少鹏. 一种有效的差分隐私事务数据发布策略 [J]. 计算机研究与发展, 2014, 51 (10): 2195-2205.

[2] 欧阳佳，印鉴，刘少鹏. 一种分布式事务数据的差分隐私发布策略 [J]. 软件学报, 2015, 26 (6): 1457-1472.

[3] Goethals B, Laur S, Lipmaa H, et al. On private scalar product computation for privacy-preserving data mining [C]// Proc of the 7th International Conference on Information Security and Cryptology. 2004: 104-120.

[4] Aggarwal C C, Philip S Y. A general survey of privacy-preserving data mining models and algorithms [M]. [S. l.] : Springer, 2008.

[5] Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis [C]// Proc of the 3rd Conference on Theory of Cryptography Theory of Cryptography. 2006: 265-284.

[6] Dwork C. Differential privacy [C]// Proc of International Colloquium on Automata, Languages and Programming. 2006: 1-12.

[7] Dwork C. Differential privacy in new settings [C]// Proc of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics. 2010: 174-183.

[8] Dwork C. Differential privacy: a survey of results [C]// Proc of the 5th Conference on Theory and Applications of Models of Computation. 2008: 1-19.

[9] Lee J, Clifton C. Differential identifiability [C]// Proc of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data mining. [S. l.] : ACM Press, 2012. 1041-1049.

[10] Evfimievski A, Gehrke J, Srikant R. Limiting privacy breaches in privacy preserving data mining [C]// Proc of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. [S. l.] : ACM Press, 2003: 211-222.

[11] Sweeney L. k-anonymity: a model for protecting privacy [J]. International Journal of Uncertainty Fuzziness and Knowledge Based Systems. 2002, 10 (5): 557-570.

[12] [Machanavajjhala A, Kifer D, Gehrke J, et al. l-diversity: Privacy beyond k-anonymity [J]. ACM Trans on Knowledge Discovery from Data, 2007, 1 (1): 1-12.